

## DESCRIPTION

## IMAGE COMPOSITION APPARATUS AND METHOD

## 5 TECHNICAL FIELD

The present invention relates to an image composition apparatus that displays a real image superimposed with another image on a display unit of display means to be worn on a head.

10

## BACKGROUND ART

Conventionally, upon shooting movie or television program scenes, a performer acts according to memorized script contents. After shooting one scene, a director  
15 gives directions about that scene to the performer, the performer confirms the directions while observing playback of a video of himself or herself, and shooting progresses while reflecting those directions in action. In such process, shooting is made.

20 However, it is a heavy burden for a performer to memorize the script contents. Since the director gives directions after shooting one scene, the performer cannot receive fine directions from the director during action. Also, the performer cannot see a video of himself or herself,  
25 i.e., how he or she is acting, during shooting.

## DISCLOSURE OF INVENTION

It is an object of the present invention to provide an image composition apparatus which allows a user to observe a real image superimposed with another image while wearing a display on his or her head. Note that the real  
5 image may be either an image which is taken by a video camera provided to the display that the user wears on the head, and is displayed on the display (to be referred to as video see-through hereinafter), or a real space observed via the display (to be referred to as optical see-through  
10 hereinafter).

It is another object of the present invention to display another image to be superimposed at a position that does not disturb observation of a real image.

It is still another object of the present invention  
15 to turn on/off superimposed display and to switch displayed contents by a predetermined gesture of a person who wears the display, and to improve operability.

Other features and advantages of the present invention will be apparent from the following description  
20 taken in conjunction with the accompanying drawings, in which like reference characters designate the same or similar parts throughout the figures thereof.

#### BRIEF DESCRIPTION OF DRAWINGS

25 The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate embodiments of the invention and, together with the

description, serve to explain the principles of the invention.

Fig. 1 shows a use example of an image composition apparatus according to an embodiment of the present  
5 invention;

Fig. 2 is a block diagram showing the use example of Fig. 1;

Fig. 3A is a perspective view of an HMD (Head Mount Display) in Figs. 1 and 2 when viewed from the front side;

10 Fig. 3B is a perspective view of the HMD (Head Mount Display) in Figs. 1 and 2 when viewed from the rear side;

Fig. 4 shows the generation processes of a video to be superimposed on the HMD;

Fig. 5 is a diagram showing the configuration of  
15 programs which run on an information processing apparatus 300 shown in Fig. 2;

Fig. 6 is a flow chart showing the process of an HMD display thread 1000 shown in Fig. 5;

Fig. 7 is a flow chart showing the process for  
20 determining a display position in step S103 in Fig. 6;

Fig. 8 is a flow chart showing the process of a terminal management thread 2000 in Fig. 5;

Fig. 9 is a flow chart showing the process of a script management thread 3000 in Fig. 5; and

25 Fig. 10 is a state transition chart showing transition of the state of a gesture recognition thread 4000.

## BEST MODE OF CARRYING OUT THE INVENTION

Preferred embodiments of the present invention will now be described in detail in accordance with the  
5 accompanying drawings.

Fig. 1 shows an example of the system arrangement of an image composition apparatus according to an embodiment of the present invention. Fig. 2 is a block diagram showing the system arrangement in Fig. 1. Figs. 3A and 3B are  
10 perspective views showing an HMD (Head Mount Display) in Figs. 1 and 2, in which Fig. 3A is a perspective view of the HMD when viewed from the front side, and Fig. 3B is a perspective view of the HMD when viewed from the rear side.

Reference numeral 100 denotes an HMD (Head Mount  
15 Display) a person wears on his or her head; 200, a three-dimensional position sensor main body; 210, a three-dimensional position sensor fixed station; 220, a hand position sensor; 300, an information processing apparatus; 400, a video camera; 500, a video deck; and 600,  
20 a terminal. These components constitute an image composition apparatus.

The HMD 100 is an eyeglass-type image display apparatus that adopts a video see-through system. This HMD  
100 comprises a right-eye camera 110, left-eye camera 111,  
25 HMD built-in sensor (three-dimensional sensor mobile station) 120, right-eye LCD (Liquid crystal display) 130, and left-eye LCD 131. Note that the HMD 100 may adopt a

combination of an optical see-through HMD and video camera in place of the video see-through HMD.

The right-eye camera 110 of the HMD 100 is connected to a video capture card 320 of the information processing apparatus 300, and the left-eye camera 111 is connected to a video capture card 321 of the information processing apparatus 300. The right-eye LCD 130 is connected to a video card 330 of the information processing apparatus 300, and the left-eye LCD 131 is connected to a video card 331 of the information processing apparatus 300. The LCDs 130 and 131 display a composite video of those actually captured by the left-eye camera 111 and right-eye camera 110, and, for example, script data or a video from the video camera 400 (Fig. 4).

The video to be displayed on the HMD 100 is generated by the information processing apparatus 300. The information processing apparatus 300 comprises a CPU 301, memory 302, PCI bridge 303, hard disk I/F 340, hard disk 350, and the like in addition to a serial I/O 310, the video capture cards 320, 321, and 322, and video cards 330 and 331 mentioned above.

The three-dimensional position sensor 200 comprises the three-dimensional position sensor fixed station 210 and the three-dimensional sensor mobile station 120 which is built in the HMD. The three-dimensional position sensor 200 measures the relative position between the three-dimensional position sensor fixed station 210 and

three-dimensional sensor mobile station 120 by magnetism. Position information has six degrees of freedom: X, Y, Z, ROLL, PITCH, and YAW. The three-dimensional position sensor main body 200 communicates with the information processing apparatus 300 via a serial interface. The position of the three-dimensional position sensor fixed station 210 is precisely measured in advance, and the absolute position (to have the center of a studio as an origin) of the HMD can be detected by detecting the relative position of the three-dimensional sensor mobile station 120.

The three-dimensional sensor mobile station 120 is connected to the three-dimensional position sensor main body 200. The three-dimensional position sensor fixed station 210 is connected to this three-dimensional position sensor main body 200, and the hand position sensor 220 that the person who wears the HMD 100 wears on the hand is further connected. The hand position sensor 220 has the same structure as the three-dimensional sensor mobile station 120, and is also connected to the three-dimensional position sensor main body 200. In this embodiment, the hand position sensor 220 and three-dimensional position sensor 200 communicate with each other by magnetism. The three-dimensional position sensor main body 200 is further connected to the serial I/O 310 of the information processing apparatus 300. These sensors measure both the posture of a person who wears the HMD 100, and the position of the HMD 100.

This embodiment uses a magnetic position sensor which is separated into a fixed station and position sensor, but may use a position sensor using a gyro as long as it can measure the position of the HMD. The terminal 600 is used  
5 to input instructions from a staff member or to input shooting start and stop instructions.

The configuration of programs which run on the information processing apparatus 300 will be explained below. In the following description, assume that a  
10 performer wears the HMD 100 in rehearsal upon shooting a movie or television program.

Fig. 5 shows the configuration of programs which run on the information processing apparatus 300 in Fig. 2. The programs include an HMD display thread 1000, terminal  
15 management thread 2000, script management thread 3000, and gesture recognition thread 4000. Data are exchanged among the threads via an instruction buffer 2001, script buffer 3002, and display mode flag 4001.

The HMD display thread 1000 displays videos captured  
20 by the right-eye camera 110 and left-eye camera 111 on the LCDs 130 and 131. In this case, the thread 1000 superimposes an instruction from a staff member written in the instruction buffer 2001 or script data written in the script buffer 3002. Also, an image taken by the television  
25 camera 400 is also superimposed.

The process of the HMD display thread 1000 will be explained below. Fig. 6 is a flow chart showing the process of the HMD display thread 1000 in Fig. 5.

Note that the processes that pertain to the eyes of a person who wears the HMD 100 have no difference for the right and left eyes, and are basically the same. Hence, the process for the right eye will be explained below.

A video from the video camera 100 is captured by the information processing apparatus 300 via the video capture card 320 (step S100). The captured video is written on a video buffer on the memory 302 (step S101). This video buffer is a work area for storing a video during generation. By writing the captured actually taken video, an actually taken video can be used as a background image.

It is checked if an information display mode for displaying information such as a video from the video camera 400, a script, an instruction from a staff member, or the like is set. If the information display mode is set, an information display process starts (step S102).

In the information display process, the display position of information is determined to be a position where information does not disturb the performer (step S103). In this embodiment, the display position of information is set to be a position where information overlaps a table, the coordinate position of which is known, so that the information does not occlude another performer. Note that



determination method of the information display position will be explained later.

After the information display position is determined, a video from the video camera 400, script data 350, and an instruction from a staff member via the terminal 600 are captured (step S104), and the captured information is written in a video buffer area corresponding to the determined display position (step S105). Since the video from the HMD 100 has already been written in the video buffer, information is superimposed on that video.

Upon completion of rendering, the contents of the video buffer are transferred to a frame buffer on the video board 330 to display (render) the contents (video) of the video buffer on the LCD 130 in the HMD 100 (step S106).

The determination method of the information display position in step S103 will be described below.

Fig. 7 is a flow chart of the process for determining the information display position in step S103 in Fig. 6.

The position of the HMD 100 is acquired from the three-dimensional sensor main body 200 (step S200). The information processing apparatus 300 generates and sets a modeling conversion matrix on the basis of the position acquired in step S200, the coordinate position of the HMD 100, and parameters such as the focal length of the camera and the like, which are measured in advance, so as to obtain an image from the viewpoint of the HMD 100 (step S201). Note that the "position and posture of the viewpoint of an

observer", which are required to estimate a real image that a person who wears the HMD 100 observes, are detected using the HMD built-in position sensor 120 in this embodiment, but the present invention is not limited to such specific  
5 method. A position sensor separate from the HMD 100 may be attached to the head of the person to detect the position and posture of the head.

The generated modeling conversion matrix converts the known coordinates of the four corners of the table into  
10 points on the screen of the LCD 130 (step S202). With this conversion, the display positions of the four corners of the table within the screen of the HMD 100 can be determined.

It is checked if all the four corners fall within the screen (step S203). If all the four corners fall within  
15 the screen (YES in step S203), the center of the four corners is set to be the information display position (step S204). In this way, information is displayed on the table. If the four corners do not fall within the screen (NO in step S203), a predetermined lower right position is set to be the  
20 central position of information display (step S205).

An example of the aforementioned image composition process will be explained below with reference to Fig. 4. Fig. 4 shows a state wherein a video 4b taken by the video camera 400, and information from the script data 350 are  
25 superimposed on a video 4a obtained by the video cameras 110 and 111 of the HMD 100. With the process in step S103 mentioned above, the display position of the video from the

video camera 400 and information from the script data 350 are determined to be the center of the four corners of the table, and the video and the information are rendered, as indicated by a video 4c. The videos 4a and 4c are composited  
5 to obtain a video 4d, which is observed by the person who wears the HMD 100. Note that Fig. 4 typically illustrates the respective videos, and some of the video contents, composite positions, and the like are not accurate.

In this embodiment, as described previously, since  
10 the information is superimposed on the table, the coordinate position of which is known, the field of view of the performer can be prevented from being intercepted. Display of information on the table does not limit the present invention. For example, information may be  
15 superimposed on a portion of a wall, the coordinate position of which is known. Furthermore, the position of the display can be dynamically changed. For example, the position of the display may be changed from a desk to a wall.

The terminal management thread 2000 in Fig. 5 mainly  
20 processes an input from the terminal 600, and writes an instruction from a staff member to the performer via the terminal 600 in the instruction buffer 2001. At the same time, the terminal management thread 2000 informs the script management thread 3000 of staff member's operations.

25 The process of the terminal management thread 2000 will be described below.

Fig. 8 is a flow chart showing the process of the terminal management thread 2000 in Fig. 5.

The terminal management thread 2000 normally waits for an input from the terminal 600 (step S300). In this embodiment, a shooting start or stop command or the like is issued by inputting a character string to the terminal 600. However, the present invention is not limited to such specific instruction method to the terminal 600, and various other known user interfaces may be used.

10        It is checked if an input character string is a shooting start command (step S301). If the input character string is a shooting start command, a script output thread starts (step S304). If it is determined as a result of checking in step S301 that the input character string is not a shooting start command, it is checked if the input character string is a shooting stop command (step S302). If the input character string is a shooting stop command, the script output thread stops (step S305). If it is determined as a result of checking in step S302 that the input character string is not a shooting stop command, it is determined that the input character string is an instruction from a staff member to the performer, and that character string is written in the instruction buffer 2001 (step S303). After that, the flow returns to step S300.

25        The script management thread 3000 writes script data stored in the hard disk 350 in the script buffer 3002 at timings according to time stamps stored in the script data.

The script data is stored as a sequence of sets of time stamps and character strings to be displayed at those timings.

The process of the script management thread 3000 will be described below.

5           Fig. 9 is a flow chart showing the process of the script management thread 3000 in Fig. 5.

As an initial setup process, an internal shooting clock 3001 is reset to zero, and a pointer of script data is returned to the head of the script (step S400).

10           Data (next script data to be displayed) pointed by the pointer of the script data is loaded from the hard disk 350 (step S401). The control waits until the shooting clock 3001 becomes the same as the time stamp of the script data pointed by the pointer (step S402).

15           The script data pointed by the pointer is written in the script buffer (step S403). The pointer of the script data is advanced (step S404), and the flow returns to step S401 to repeat the aforementioned steps.

          The gesture recognition thread 4000 in Fig. 5  
20 recognizes a gesture (hand position and posture) of the performer on the basis of the position of the hand position sensor 220 obtained via the three-dimensional position sensor main body 200. Every time a gesture is recognized, the display mode flag 4001 is turned on/off.

25           In this embodiment, as a gesture for turning on/off display, an action for moving the hand up and down three times for a second is selected.

Fig. 10 is a state transition chart showing transition of the state of the gesture recognition thread 4000.

The gesture recognition thread 4000 is normally in  
5 a "standby state S500". At this time, an internal counter for counting upward and downward actions of the hand is cleared to zero. Upon detecting an upward acceleration from the hand position sensor 220, the gesture recognition thread 4000 transits to an "upward acceleration state  
10 S501".

When the acceleration has stopped, the gesture recognition thread 4000 transits to an "upward acceleration stop state S502". However, if an upward acceleration is detected again within 0.1 sec after transition, the gesture  
15 recognition thread 4000 returns to the "upward acceleration state S501".

If no acceleration is detected within 0.1 sec after transition to the "upward acceleration stop state S502", it is determined that the action detected is merely an  
20 upward movement of the hand but is not a gesture, and the gesture recognition thread 4000 transits to the "standby state S500".

When an upward acceleration is detected in the "upward acceleration stop state S502", the gesture  
25 recognition thread 4000 transits to the "upward acceleration state S501".

When a downward acceleration is detected in the "upward acceleration stop state S502", the gesture recognition thread 4000 transits to a "downward acceleration state S503". When the acceleration has  
5 stopped, the gesture recognition thread 4000 transits to a "downward acceleration stop state S504". When a downward acceleration is detected again within 0.1 sec after transition, the gesture recognition thread 4000 returns to the "downward acceleration state S503".

10 When an upward acceleration is detected in the "downward acceleration stop state S504", the gesture recognition thread 4000 transits to the "upward acceleration state S501". At this time, the internal counter is incremented. This corresponds to a case wherein  
15 the hand is moved downward after upward movement.

When no acceleration is detected within 0.1 sec after transition to the "downward acceleration stop state S504", the gesture recognition thread 4000 transits to the "standby state S500". If counter = 3, it is determined  
20 that the gesture is complete, and an event is generated to invert the value (TRUE/FALSE) of the display mode flag 4001.

The gesture recognition thread 4000 executes the process according to the aforementioned state transition chart to detect an event. In the above description, a  
25 full-superimpose display ON/OFF instruction is issued by a gesture. However, for example, display of script data, instruction data, and video camera image may be

individually turned on/off, as is known to those who are skilled in the art.

As described above, according to the image composition apparatus of this embodiment, various kinds of  
5 information can be given to the performer by superimposing a desired image.

Such information can be displayed at a position (e.g., on the wall of a shooting background) where the displayed information does not disturb the field of view of the  
10 performer.

In this way, the performer can confirm lines, director's instruction, taken video, and the like without largely moving the line of sight, and the load on the performer can be reduced.

15 The performer can act while observing a video which is being taken by a cameraman, and which cannot so far be confirmed until after action.

Furthermore, since display of such information can be turned on/off by a performer's gesture, information can  
20 be easily displayed at a desired timing in rehearsal.

Note that the objects of the present invention are also achieved by supplying a storage medium, which records a program code of a software program that can implement the functions of the above-mentioned embodiments to the system  
25 or apparatus, and reading out and executing the program code stored in the storage medium by a computer (or a CPU or MPU) of the system or apparatus.



In this case, the program code itself read out from the storage medium implements the functions of the above-mentioned embodiments, and the storage medium which stores the program code constitutes the present invention.

5       As the storage medium for supplying the program code, for example, a floppy disk, hard disk, optical disk, magneto-optical disk, CD-ROM, CD-R, magnetic tape, nonvolatile memory card, ROM, and the like may be used.

10       The functions of the above-mentioned embodiments may be implemented not only by executing the readout program code by the computer but also by some or all of actual processing operations executed by an OS (operating system) running on the computer on the basis of an instruction of the program code.

15       Furthermore, the functions of the above-mentioned embodiments may be implemented by some or all of actual processing operations executed by a CPU or the like arranged in a function extension board or a function extension unit, which is inserted in or connected to the computer, after  
20       the program code read out from the storage medium is written in a memory of the extension board or unit.

25       As described in detail above, according to the image composition apparatus, since another image is displayed on a display unit that displays a real image, the other image can be superimposed on the real image. Hence, the user who wears display means on the head can observe the other image superimposed on the real image.

When text information is displayed as the other image, a text instruction can be given to the user who wears the display means on the head.

5       The other image can be an image from a viewpoint other than that of the user who wears the display means, which is taken by image taking means (television camera 400). For this reason, when this image taking means takes an image of the user who wears the display means on the head, the user can confirm his or her actions in real time.

10       Superimpose display of the other image in the HMD 100 is turned on/off on the basis of the hand position, hand movement, posture, and the like of the user who wears the display means. For this reason, the user can easily turn on/off superimpose display of the other image according to  
15 his or her will by changing the hand position and posture.

      Since the superimpose display region of the other image is determined based on the posture of the user who wears the display means, the other image can be displayed at a position where that image does not disturb the user.

20       As many apparently widely different embodiments of the present invention can be made without departing from the spirit and scope thereof, it is to be understood that the invention is not limited to the specific embodiments thereof except as defined in the claims.